QRMS Project 5

Due Thursday, May 1, 2008

Please do not consult outside sources or people outside your group on this project.

Widgets, Inc. has a very complicated manufacturing process involving many machines. Central to this operation is the Widgetometer machine, which takes many days to repair if it should break down, necessitating the closure of the entire plant for this period. Thus, Widgets, Inc. tries to do preventive maintenance to keep this machine running smoothly. However, since this maintenance is itself quite expensive, the company would like to perform it only when it's "likely" that the machine will break soon. For this reason, they hired two teams of statisticians (working separately) to come up with algorithms (taking many factors which needn't concern you into account) which will provide a probability (between 0% and 100%) of the machine breaking during the next month, with 0% meaning it definitely won't and 100% meaning it definitely will. (Widgets, Inc. tries to do maintenance if the probability of breaking is at least 50%, although the exactly percent needed to trigger repairs changes from month to month depending on their cash flow.) Since sales projections are also important, Widgets, Inc. also asked both teams to create a second algorithm to predict the number of widgets that the factory will produce in the upcoming month. Additionally, Old Gus (who has been working at the plant since it opened in 1962) thinks that he can predict the probability of the machine breaking and the number of widgets that will be produced based on his incredible intuition. Unfortunately, this has caused a great deal of confusion for the management of Widgets, Inc., since they're not sure which of the three recommendations to use in any given month! Although they've compared the algorithms predictions against their historical records (see the data sheet on the last page), they don't know how to determine which algorithms are the most accurate overall, which is where you come in: in a well-organized typed essay of 3-5 pages, working alone or in a group of up to 3 do the following:

(1) provide a method for providing a quantitative rating (that is, one or more numbers) for each algorithm based on its past history. Note that you'll be providing two different methods: one for rating the algorithms that predict the probability of machine failure, and one for rating the algorithms that predict the number of widgets that will be produced
(2) explain why your methods will provide a reasonable measure of the "accuracy" of each algorithm and why any assumptions you made in creating the method are justified
(3) discuss the strengths and weaknesses of your methods
(4) calculate what rating your algorithm gives for each of the three algorithms and determine which one(s) the company should use in the future
(5) justify why the algorithms (one for each of the two questions) you are recommending are the best

Note you should answer the above questions separately for the two different methods (that is, even if you think that Algorithm #1 is best for predicting machine failure, you're still free to say that Algorithm #3 is best for predicting widget production, for example). Also, Widgets, Inc. would like to point out that having the Widgetometer break down is much more costly in the long run than performing occasional unnecessary maintenance on it, so it's worse for an algorithm to say that maintenance is unnecessary when it's actually needed than it is for an algorithm to say that maintenance is necessary when in fact it isn't (as long as this doesn't happen too often), so your method should take this into account.

The sample data is arranged as follows: each row represents one month's data from the company's historical records. The first two columns give what actually happened in that month (i.e., the number of widgets the machines actually produced and whether the Widgetometer actually broke down ("yes" or "no"), while the remaining columns provide the predictions of the two algorithms and of Old Gus (i.e., the number of widgets that the algorithm predicted would be produced, and the predicted probability of the Widgetometer breaking (on a scale of 0 to 100)).

Although you are free to use any organizational strategy, the following outline may help you to express your ideas clearly and to ensure that you are answering all of the questions asked:

I. Introduction: Describe what problems you are trying to solve and the answers you found.
II. Assumptions: Describe any assumptions you made. If you made no assumptions, you don't need to include this. But think about it carefully: you probably made assumptions even if you don't immediately realize it, and I'll notice them even if you don't and grade accordingly.
III. Methodology: Describe your method for evaluating the data given. Remember that you need to describe *both* of your methods.
IV. Strengths and Weaknesses: Describe what's good about your methods and what's potentially bad about them. In what circumstances is it bad? Are these circumstances common?
V. Data: State how each algorithm fared under your methods. If you're using more than one number to quantify the accuracy of an algorithm, state how you used each number in order to reach your final conclusion of which was best. Make sure that you also state your final conclusions of which algorithms are the best.
VI. Interpretation: What does the data and your final answer mean? Are the numbers surprising? Larger than you expected? Smaller? Should we care? Why or why not?
VII. Further Research: Discuss related questions that you could look into. For example, would further data allow you to refine your method? Is there information that would have been helpful that wasn't given? If you were to spend a year on this problem, what would your next steps be?
VIII. Conclusion: Any final thoughts relevant to the topic? Personal reflections are acceptable here (but not required), if you want to talk about what you did or didn't learn, enjoy, etc. about this.

In the process of answering these questions, the key first step is figuring out a method for approaching the problem quantitatively. A recurring theme is this class has been to move from qualitative descriptions like "low variation" to a quantitative one like the five-number summary and the standard deviation, or from the subjective measure of "statistical significance" to the consideration of the "0.05 level" and "0.01 level." While none of the methods we have developed will work directly, you may be able to manipulate or combine some of the ideas we've seen in Chapters 5 through 7; however, make sure you're doing something that makes sense. For example, you *could* calculate the standard deviation of the numbers for each algorithm, but doing this would not be meaningful, since it would only measure how each algorithm deviated from its mean, which is completely disconnected from the real question of how each algorithm deviates from the actual data. Similarly, you *could* just calculate the average prediction of each algorithm and compare it to the average of the actual data, but again this would only show that the averages were the same and not that the algorithm was actually reasonably correct most of the time. Since the widget-production algorithm problem deals with the distribution of numbers, you probably want to find a way of measuring how far each algorithm's predictions differ from the actual data, so your best strategy is probably to modify a technique from Chapters 5 and 6. Since the machine break-down algorithm problem deals with probabilities, your best strategy is probably to modify a technique from Chapter 7. Of course, in both cases you're free to use any method, provided that you are able to justify its use.

Note that one of the difficulties with the machine break-down question is that in each given month the machine either breaks down or it doesn't, so you have to come up with a way to compare a number between 0 and 100 with a "yes" or a "no." Since the company may choose to base maintenance on how likely failure is, you do have to come up with a way of working with this numerically (which is to say, don't just convert the percentage data into a column of yeses and noes; while both 51% and 80% suggest the machine is more likely to break than not, they are nonetheless expressing different degrees of certainty and so are *not* the same thing: it's better for an algorithm to be confident in its predictions, and your method should take this into account.)

Helpful Hints:

(1) As in all projects: spell-check, grammar-check, proof-read. Outline your argument before you begin and give your essay a title. Staple your pages together with the rubric on front as a cover sheet: don't just do origami with the corner; it won't work. Follow standard rules of good writing and good English.

(2) Remember, you are writing on a formal, scientific topic. The rules are a bit different than colloquial writing. First, don't use colloquialisms. Second, you should focus on what your methods and conclusions are and not as much on the process by which you arrived at them.

(3) I am providing you with the grading rubric I'll use to grade this project. Look at the questions on it and make sure that you address each one.

(4) Make sure your method is a good one before you start working with the numbers; there are a lot of numbers and you don't want to waste time computing something until you're sure it will help you answer the question. If you want advice on whether your method is good, ask me in office hours or e-mail me a short description of it.

(5) Try to avoid making computations by hand: they'll take you a long time and will be error-prone. If you know Excel or another package (or want to learn it), you can use it. If not, try searching for a calculator online that can do it (which is an exception to the "no outside sources" rule for this project). For example, searching for "standard deviation calculator" will yield results such as `http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/Descriptive.htm` (which also calculates the five-number summary and other information). In particular, please don't include the details of any calculations in your paper: you just need to explain how they were done (for example, you can say "add up all of the numbers and divide by 10" without actually showing this calculation). If you used any software, online calculator, or other tool to do the calculations, you should mention what it is or cite it.

(6) Remember: you don't have to pick the same algorithm for both questions.

(7) You should come up with a quantitative rating for each algorithm separately and only compare them at the end. Although you might be tempted to do some type of system based on awarding a "point" to the algorithm which was most accurate in each month, this isn't a good idea, since the worst of the three algorithms may "steal" enough points to throw the overall outcome. (This is similar to the "third-party candidate effect" in voting.)

(8) It was mentioned above that the widget creation process is a complicated one composed of many factors, which as we've seen is the kind of thing that often gives rise to a normal distribution. You might try to find something normally distributed in this data (there is something, in fact, but it's not immediately obvious), as this is a good (but not the only or necessarily best) way of approaching the widget-production question.

(9) The data from Old Gus was generated completely at random; it's an example of what mathematicians refer to as a 'sanity check.' If your method suggests that Old Gus made the best predictions, then there's probably a flaw in your method.

(10) Many people assume mathematics is about getting the right answer. That's certainly an aspect, but in many cases (such as this one) it's unclear what the correct answer is. In cases like these, mathematics also has much to do with communication: if you can't convince others that your answer is correct, it's wrong by default.

(11) Don't wait until the last minute. This isn't a project that can be dashed off in an evening. You may hand it in early, if you wish.

(12) Have fun with this. Mathematics is supposed to be fun.